DOCUMENT RESUME

ED 173 424 TM 009 492

AUTHOR	Anderson, Lorin W.
TITLE	Handbook for the Release of Test Data. The Office of Research Report Series.
INSTITUTION	South Carolina State Dept. of Education, Columbia. Office of Research.
PUB DATE	Apr 79
NOTE	48p.; Best copy available
AVAILABLE FROM	Educational Products Center, 1206 Rutledge Building, Columbia, South Carolina 29201 (free while supply lasts)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Elementary Secondary Education; *Information Dissemination; *News Media; Norm Referenced Tests; *Parents; Parent School Relationship; *Publicize; School Community Relationship; Scores; Statistical Data; Teacher Administrator Relationship; Technical Reports; *Testing Programs; Test Interpretation; *Test Results
IDENTIFIERS	Comprehensive Tests of Basic Skills

ABSTRACT

Several questions are frequently asked by parents and news media regarding the purpose, content, and interpretation of the Comprehensive Tests of Basic Skills (CTBS)--the achievement battery currently used in South Carolina assessment programs. Proper interpretation of the CTBS, or any school district testing program, requires a sound understanding of its validity and reliability, and of the composition of the national sample relative to the students tested. It is easier to interpret test results by displaying percentile ranks, standard error of measurement, and frequency distributions in graphic form. As for releasing test data to parents and the news media, a formal report should be released at the district level, after first informing school district staff. The report should relate the purpose and results of the tests to school/district objectives, describe the students tested relative to the national sample, and compare results by year and by school. At the beginning of the school year, parents should be informed of the testing program, and results should be explained to individual parents in a parent-teacher conference. Followup meetings and press conferences are recommended for discussing programs to improve achievement. (CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
E. L. Knight

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

ED173424

TM009 492

HANDBOOK FOR THE RELEASE OF TEST DATA

Division of Administration and Planning

Robert R. Hill

Deputy Superintendent

Office of Research

Paul D. Sandifer

Director

South Carolina Department of Education

Charlie G. Williams

State Superintendent

Columbia, South Carolina

April, 1979

THE OFFICE OF RESEARCH REPORT SERIES

HANDBOOK FOR THE RELEASE OF TEST DATA

Developed for the

South Carolina Department of Education

by

Dr. Lorin W. Anderson

Associate Professor

University of South Carolina

TABLE OF CONTENTS

Page

LIST OF TABLES..... v

LIST OF FIGURES..... vi

INTRODUCTION..... 1

QUESTIONS TO ANTICIPATE AND ANSWERS TO PREPARE..... 2

What Is the Purpose of the Comprehensive Tests of Basic Skills (CTBS)?..... 2

What Type of Tests Are the CTBS?..... 3

What Do the CTBS Test?..... 3

Does What Is Measured by the CTBS Match What Is Taught in the School or District?..... 5

How Accurate Are the CTBS Scores?..... 7

What Student Characteristics Should Be Considered When Making Comparisons to the National Sample?..... 10

What Scores Are Appropriate for Reporting the Results?..... 13

Summary..... 17

SUGGESTED PROCEDURES FOR RELEASING TEST DATA..... 17

Preparing the Written Document..... 18

Informing School and District Staff..... 18

Informing the News Media..... 20

Informing Parents..... 22

Providing Follow-Ups..... 23

TABLE OF CONTENTS

(Continued)

Page

REFERENCES..... 26

APPENDIX

A. An Example of a District News Release..... 28

LIST OF TABLES

Table — Page

1. Correspondence of Test Levels with Grade Levels..... 4

2. General Skill Areas, Number of Items per Area, and Administration Times for Test Levels C, 1, 2, and A..... 6

3. Approximate Standard Errors of Measurement for Each Test by Test Level (in items and percent error)..... 9

4. Items That Should Be Contained in the Written Report Prepared by the School District. 19

LIST OF FIGURES

Figure		Page
1.	Interrelationships among General Skill Area, Subskill Areas, and Content Categories	31
2.	Interrelationships among General Skill Area, Subskill Areas, and Content Categories for Reading	32
3.	Interrelationships among General Skill Area, Subskill Areas, and Content Categories for Language	33
4.	Interrelationships among General Skill Area, Subskill Areas, and Content Categories for Mathematics	34
5.	Interrelationships among General Skill Area, Subskill Areas, and Content Categories for Reference Skills	35
6.	Interrelationships among General Skill Area, Subskill Areas, and Content Categories for Science	36
7.	Interrelationships among General Skill Area, Subskill Areas, and Content Categories for Social Studies	37
8.	A Summary of School/District Performance on the CTBS/S (by General Skill Area)	38
9.	A Summary of School/District Performance on the CTBS/S (emphasizing errors of measurement) (by General Skill Area)	39
10.	Graphical Display of School/District Results on the CTBS/S (by General Skill Area)	40

INTRODUCTION

The South Carolina Education Finance Act of 1977 contains several implications for schools and school districts. Among these are implications for the gathering and dissemination of test data. The responsibility for gathering the test data belongs to the State Department of Education. The Act requires that these data be gathered as part of the Department's Statewide Testing Program and that each district must participate in this program.

The responsibility for the dissemination of the test results belongs to each school district. As the law specifically states, "Each school district shall annually submit to the State Board of Education and to the people of the district that district's programmatic report including results of the required testing program, the annual long-range plan and the evaluation of program effectiveness by August fifteenth of each year." (Act 163 of 1977, Section 8-4-f).

Because the dissemination of test results is new for many districts, guidance concerning this dissemination may be needed. The purpose of this handbook is to provide district personnel with information that can be used to better communicate the results of the testing program to the people of the district. In order to accomplish this purpose effectively and efficiently, the handbook is organized into two general sections. The first section, "Questions to Anticipate and Answers to Prepare", poses several questions which school personnel may be asked by the news media, parents, and the general public. This section also presents information which will assist school personnel in answering these questions. The questions and

answers deal with issues specific to the Comprehensive Tests of Basic Skills, Form S (CTBS/S). The CTBS/S are the tests currently used by the State Department of Education as part of the Statewide Testing Program. The second section, "Suggested Procedures for Releasing Test Data", presents information concerning the presentation and release of test results to the news media, parents, and the general public.

This handbook was developed in order to facilitate communication of the results obtained from the required Statewide Testing Program. However, the suggestions on the presentation and release of test scores are equally applicable to information gathered from other districtwide testing programs; information that may be subject to release under the provisions of the Freedom of Information Act.

QUESTIONS TO ANTICIPATE AND ANSWERS TO PREPARE

What Is the Purpose of the Comprehensive Tests of Basic Skills (CTBS)?

As indicated in the "Test Coordinator's Handbook" (1974), the Comprehensive Tests of Basic Skills (CTBS) were "designed to measure the extent to which individual students have developed the capabilities and learned the skills that are prerequisite to studying and learning in subject matter courses, and necessary for functioning in a society based on daily use of language and number" (p. 13). Thus, the CTBS were designed to gather information concerning the extent to which students had acquired general skills (and specifically, general communication skills) which would benefit them in a wide variety of school and non-school situations. This focus on measuring learning which results from a wide range of experiences

is best stated by the developers of the CTBS: "The measurement of these skills and abilities cannot be divorced entirely from the measurement of knowledge acquired through schooling, but it is not the intent of these tests to measure this knowledge directly" (p. 13). The handbook continues, "the emphasis in this series [of tests] is on the measurement of the grasp of broad concepts and abstractions as developed by all curriculums, and on facility in the skills that are required in the effective use of language and number such as classifying, manipulating, and interpreting" (p. 13).

What Type of Tests Are the CTBS?

The CTBS are classified as "norm-referenced tests." Norm-referenced tests provide what may be termed "comparative information". That is to say, students' scores on such tests are given meaning by comparing them with the scores of a large, national sample of students. The national sample of students must be at the same grade level as the district students but may differ in terms of such characteristics as socio-economic status. Thus, the test results from norm-referenced tests indicate how the general skills possessed by students within a given school or district compare with those possessed by a national sample of students.

What Do the CTBS Test?

There are seven levels of the CTBS which are appropriate for students of different ages and enrolled in different grades. Table 1 indicates the test levels that are appropriate for students at different grade levels.

Table 1

Correspondence of Test Levels with Grade Levels

Test Level	Grade Level
A	Beginning of Kindergarten to Beginning of First Grade
B	Mid-Kindergarten to End of First Grade
C	Mid-First Grade to End of Second Grade
1	Mid-Second Grade to End of Fourth Grade
2	Mid-Fourth Grade to End of Sixth Grade
3	Mid-Sixth Grade to End of Eighth Grade
4	Mid-Eighth Grade to End of Twelfth Grade

The Statewide Testing Program tests students at Grades 3, 6, and 11 and uses four test levels: Level C, Level 1, Level 2, and Level 4. Levels C and 1 are administered at Grade 3, Levels 1 and 2 are administered at Grade 6, and Level 4 is administered at Grade 11. The composition of these four levels is very similar. Each level (except Level C) consists of six general skill areas: reading, language, mathematics, reference skills, science, and social studies. Level C does not include a test of reference skills.

Three of the general skill areas, reading, language, and mathematics are divided into subskill areas. Reading is divided into two subskill areas: vocabulary and comprehension. Language is divided into three sub-

skill areas: spelling, language mechanics, and language expression. Mathematics is divided into two subskill areas: computation and concepts/applications. Finally, each general skill area is divided into what are called content categories. The relationship among general skill areas, subskill areas, and content categories is shown graphically in Figure 1. The specific relationships for reading, language, mathematics, reference skills, science, and social studies are shown graphically in Figures 2 through 7.

Typically, student scores are reported for each general skill area, although the reporting of scores by subskill areas or by content category is possible. Student scores are initially computed by adding up the number of items in the general skill areas answered correctly by each student. The number of items testing each of the six general skill areas on each level of the tests are displayed in Table 2. The time necessary for the administration of each set of items is also presented in Table 2.

Does What Is Measured by the CTBS Match What Is Taught in the School or District?

In addition to the organization of the tests and the number of items on the tests, the technical quality of the test should also be considered. One indicator of a test's technical quality is validity.

Validity refers to the extent to which the test measures the skills and abilities that are deemed important by the school or district. As such, validity can be examined by comparing the skills and abilities which underlie the test with the skills and abilities which make up the school's or district's curriculum or are contained in the goals of the school or

Table 2

General Skill Areas, Number of Items per Area, and Administration Times for Test Levels C, 1, 2, and 4

General Skill Area	Level C		Level 1		Level 2		Level 4	
	# of Items	Time(min.)	# of Items	Time(min.)	# of Items	Time(min.)	# of Items	Time(min.)
Reading	74	56	85	50	85	49	85	46
Language	79	55	94	48	105	49	85	49
Mathematics	53	59	98	75	98	75	98	70
Reference Skills	--	--	20	15	20	10	20	10
Science	30	35	37	40	36	40	40	40
Social Studies	30	35	37	40	37	40	39	40

district. Since the curriculum and goals of schools or districts may vary, each school or district should examine the validity of the test for that school or district.

This examination is best made at the content category level of the tests. The content categories have been mentioned in Figures 2 through 7 and are described in greater detail in the Test Coordinator's Handbook on Pages 33 - 58. A panel of teachers should examine the content categories underlying the tests at each grade level. For each content category, consensus should be reached concerning whether that content category is taught at that particular grade level.

The validity of the tests can be reported in terms of the percent of the content categories on the test which are included in the curriculum at each grade level. For example, the language area is composed of eight content categories. If five of the eight categories are included in the sixth grade language arts curriculum, the validity of the test can be said to be 5/8ths or 62 1/2 percent. The higher this figure is, the more valid the test is with respect to the curriculum. If this figure falls much below 50 percent, one should question the applicability or validity of the test for the curriculum. Consequently, the meaning of the test results in terms of students' learning of the curriculum is unclear in this instance.

How Accurate Are the CTBS Scores?

In addition to validity, a second technical characteristic which should be considered is reliability. Reliability is concerned with the amount of error that is present when one describes a person's performance on a test.

Tests are not infallible. Since tests are not perfect indicators of students' learning, some error is inevitable. Many test users, however, fail to take this unreliability or error into consideration. In testing terminology, this error is represented by the standard error of measurement. Put simply, the standard error of measurement can be thought of as the average amount of error in a student's score. For example, the average amount of error on the test of reading is 4 items. This is to say, if the student were to take the reading test on two separate occasions, the student's score might differ by as much as 4 items. Measurement specialists take this error into consideration by subtracting it from and adding it to a student's score. Thus, a student who answers 70 items correctly on the reading test would likely have a true reading score somewhere between 66 (70 - 4 items) and 74 (70 + 4 items). Thus, the standard error of measurement is important in order to understand the precision that can be attributed to a test score.

The developers of the CTBS have computed the standard error of measurement for each test at each level. The standard errors of measurement are presented in Table 3.

As can be seen in Table 3, the errors range from 3 items to 4 items and are highly consistent over all four levels of the tests used in the Statewide Testing Program. If this error is divided by the number of items which measure a given general skill area, the percent of error on each test can be computed. When this is done, the "most dependable" tests (e.g., Level 1, mathematics) have errors of approximately 4 percent while the "least dependable" tests (e.g., Level 1, reference skills) have errors of

Table 3

Approximate Standard Errors of Measurement for Each Test by Test Level
(in items and percent error)

General Skill Area	Level C		Level 1		Level 2		Level 4	
	# of Items	% Error	# of Items	% Error	# of Items	% Error	# of Items	% Error
Reading	3	4%	4	5%	4	5%	4	5%
Language	3	4%	4	4%	4	4%	4	5%
Mathematics	3	6%	4	4%	4	4%	4	4%
Reference Skills	-	--	2	10%	2	10%	2	10%
Science	2	7%	3	8%	3	8%	3	8%
Social Studies	2	7%	3	8%	3	8%	3	8%

approximately 10 percent. These errors must be taken into consideration in order to make accurate interpretations of student performance on the tests. Since this is the case, more will be said about the relationship of standard error of measurement to test interpretation later in this handbook.

What Student Characteristics Should Be Considered When Making Comparisons to the National Sample?

Student responses to the CTBS test items are scored as correct or incorrect. The sum total of all correct responses made for each general skill area test is referred to as the raw score. Thus, if Edna answers 52 out of 94 items correctly on the language test, Edna's raw score is 52. Since little is known about the difficulty or complexity of the test items, raw scores have little meaning. Certainly 20 correct on an easy 30-item test is not the same as 20 correct on a difficult 30-item test.

For results on tests like the CTBS to be meaningful, the performance of a student or a group of students must be related to some known reference group. The most frequently used reference group is a national sample of students who are in the same grade as the students of interest. This reference group often is called a normative sample and tests like the CTBS often are termed norm-referenced tests.

In order to make a proper interpretation of a school's or district's performance, one must first compare the normative sample with the school or district sample on certain relevant characteristics. Examples of such characteristics might include age, sex, ethnic group membership, and social class. If the normative sample and the school or district sample are very

dissimilar with respect to these characteristics, then comparisons of the test performances of these two samples are not very meaningful.

Why are such comparisons not very meaningful? Consider an example. Suppose the national sample of third grade students contains 50 percent boys and 50 percent girls. Furthermore, suppose a particular district (for example, District #99) has 70 percent boys and 30 percent girls in its third grade. Because of the relatively large difference between the proportions of boys in the normative sample and in the district sample, the test results may be expected to be different for the two samples regardless of the quality of education in that particular district. Furthermore, because prior research suggests that third grade girls tend to read "better" than third grade boys, the direction of the difference in test performances can be estimated. We would expect the students in the national sample to score higher than the students in District #99 since there is a greater proportion of girls ("good readers") in the national sample than there is in the district sample.

Notice that this difference does not reflect on the quality of the education given to the students in District #99. Rather, this difference mirrors the difference in the composition of the district sample and the national sample on a relevant characteristic, namely, sex of a student. In sum, then, schools or districts that differ greatly in sex, social class, and/or racial composition from the national sample also can be expected to differ in test performance from the national sample. The direction of the difference depends on what we know from previous research.

Therefore, for proper interpretation of test scores, a comparison of

the school or district sample with that of the national sample on important characteristics is necessary. This is best accomplished by first describing the school or district in terms of such characteristics as the (1) proportion of male students, (2) proportion of Caucasian students, and (3) the proportion of students whose parents hold professional or white collar jobs. The national sample used for standardizing the CTBS also can be described in terms of these characteristics. A description of the national sample in terms of these and other relevant characteristics can be found on Pages 7 and 8 of the Technical Bulletin No. 1.

To the extent that "marked differences" exist between the students in the district and the students in the national sample on these descriptors, a comparison of the relative achievement of these two samples is potentially meaningless and should be made with caution. As has been suggested earlier, if the samples are "very different", then any differences in the test performance may be attributed to differences in the characteristics of students attending the school or district rather than to differences in the quality of education provided by the district, schools, and teachers.

In the previous paragraph phrases like "marked differences" and "very different" have been used. A question that might be asked is, "How large does this difference have to be in order to be a 'marked difference'?". This is a difficult question to answer. Certainly, a difference of 2 or 3 percent is not large enough. On the other hand, the difference does not have to reach 20 percent before the difference is worth noting. As a rule of thumb, differences of approximately 8 to 10 percent should be considered

worth noting.

What Scores Are Appropriate for Reporting the Results?

For the remainder of this discussion, let us suppose that the students in the district sample and the students in the national sample are quite similar with respect to the aforementioned descriptors. How would the comparison of test results be made?

By far the easiest and most easily understood comparison is given by the percentile rank. The percentile rank indicates the percent of students in the national sample whose test performance is equalled or surpassed by a particular student or group of students in the district sample. For example, if Alfred's test performance is equivalent to a percentile rank of 59, this number indicates that his test performance was as high or higher than 59 percent of the students in the national sample.

Because of measurement error in the test, a range of "likely" test performance would be more appropriate than a single number. Thus, if one standard error of measurement were taken into consideration, Alfred's test performance could be said to have equalled or surpassed the test performance of somewhere between 50 and 68 percent of the students in the national sample. The initial reaction to such a general interpretation of Alfred's performance on the test might be one of amazement given many people's belief in the accuracy of test results. Such an interpretation, however, is the best that can be made, given the fallibility of the test.

Percentile ranks can be used to make interpretations of schoolwide or districtwide test performance in addition to interpretations of individual

student performance. The simplest way of making an appropriate interpretation is to identify the top student, most nearly average student, and the bottom student at each grade level in the school or district. The top student would be the one having the highest total battery score. (The total battery score is based on the student's performance on the reading, language, and mathematics tests.) The most nearly average student would be the one whose total battery score is higher than one-half of the students in the school or district and lower than one-half of the students in the school or district. The bottom student would be the one whose total battery score is lower than all other students in the school or district. Given these three students, a simple summary figure can be constructed by graphing the percentile rank for each student on each general skill area. An example of such a figure is displayed in Figure 8.

Figure 8 shows at a glance how the top student in the school or district compares with the top student in the national sample. Similar comparisons are evident for the middle student and the bottom student. Summary statements can be made directly from the figure. For example, the figure indicates that our average students "do as well as" the average student in the nation in the areas of reading and science; the average students "do somewhat better than" the average student in the nation in the areas of mathematics and social studies; and they "do somewhat poorer than" the average student in the nation in the areas of language and reference skills.

In order to provide more exact information, the standard error of measurement can be indicated on the figure by short vertical lines. These

short vertical lines or bands can be copied directly from the Individual Student Record which is provided to the school or district by the State Department of Education. An example of a summary figure complete with standard error bands is shown in Figure 9. As can be seen in Figure 9, the average student in the school or district is, in fact, lower in the area of language but is not reliably lower in the area of reference skills.

Although this graphical display is probably the simplest method of displaying the data, the graph does have one drawback. The graphical display fails to take into consideration all of the students in a particular grade in the school or district. In order to display the performance of all students, a second type of figure which can provide supplementary information to the first type of display is preferable. The rationale and means for constructing this display follow.

Suppose the national sample were divided into four equal groups on the basis of total battery score. Further, suppose that these groups are called "quarters" (since there are four of them). If this were done, the top quarter (Q4) would include students with percentile ranks from 76 to 99. The upper middle quarter (Q3) would include students with percentile ranks from 51 to 75. The lower middle quarter (Q2) would include students with percentile ranks from 26 to 50, and the lowest quarter (Q1) would include students with percentile ranks from 1 to 25.

Now, by definition, one-fourth or 25 percent of the national sample would fall into each of these quarters. Thus, a comparison of a school's or district's performance with the performance of the national sample can be made by computing the number and percent of students in the school or

district that fall into each of the four quarters. Since the School Frequency Distribution computer printout is distributed to each school and the District Frequency Distribution computer printout is distributed to each district, the percentages in each quarter of the national score distribution can be taken directly from the frequency distribution report.

Once the percent of students at each grade level in the school or district who fall into each quarter is determined, the results can be summarized in a graphical form similar to that shown in Figure 10. Again, a glance at the figure provides useful information. For example, the top students in the district (that is, those in the top quarter) "do somewhat better" in reading than their national counterparts. And, the bottom students in the district (that is, those in the lowest quarter) "do more poorly" in reading than their national counterparts. The other students in the district "do just about as well" as their national counterparts. In addition, by adding up the percent of students in the two upper quarters of the national score distribution (that is, Q4 and Q3), the percent of students in the district that score above the national average can be computed quickly. For example, 55 percent of the students in the hypothetical district in Figure 10 scored above the national average in reading.

Although this second type of display does easily allow for the presentation of the performances of all students, this display does not easily allow for the presentation of standard errors of measurement. Therefore, in order to provide a more complete picture of the test results, use of the second display in conjunction with the first display is

recommended.

Summary

In sum, then, proper interpretation of test results requires a sound understanding of the validity and reliability of the test and the composition of the national sample of students relative to that of the school or district sample. To the extent that (1) the test possesses an adequate degree of validity relative to the school or district curricula in the general skill areas, (2) the error of measurement is reasonable and remembered, and (3) the composition of the national sample of students is similar to that of the school or district sample, appropriate interpretation can be made. Such interpretations can be made most easily when the test results are presented in graphic form. Three such graphic forms (Figures 8, 9, and 10) seem feasible, given the desire to improve communication with a variety of audiences.

SUGGESTED PROCEDURES FOR RELEASING TEST DATA

As the National School Public Relations Association (1976) has pointed out, "assessment programs tend to be controversial" (p. 9). The major reasons for this controversy, the association contends, seem to be (1) a lack of awareness on the part of the district personnel of the different audiences to whom the information is to be presented and (2) a lack of understanding of the meaning of various terms which are crucial for understanding test results on the part of school personnel and the members of various audiences.

The purpose of this section is to provide suggestions concerning what to tell whom and how to communicate test results in a meaningful fashion to the two major audiences, the news media and parents. In order to accomplish this purpose, this section will begin with a brief description of how to prepare a written document for releasing the test data, followed by a discussion of informing school and district staff of the test results. Special concerns for releasing tests results to the news media and parents will also be voiced. Issues dealing with ways of following up the release of test scores via parent conferences, public meetings, and the like conclude this handbook.

Preparing the Written Document

Test results should be released at the district level by the district superintendent. To aid in communication, a written document should be prepared by the district test coordinator before discussing the test results with either the news media or the public. The written document should contain the items mentioned in Table 4.

Informing School and District Staff

Once the document has been prepared, it should be distributed to school and district staff. After allowing a sufficient time for the reading of the document by school and district staff, an inservice day should be scheduled in order to discuss the test program with school and district staff. Two major purposes should be served by the inservice meeting. First, questions raised by the school and district staff should be addressed. Second, suggestions for changes in instructional programs

Table 4

Items That Should Be Contained in the Written Report Prepared by the School District

1. Purpose of the Tests (in the context of the entire educational program of the school or district)
2. Brief Description of the Tests
3. Glossary of Important Terms: Standard Error of Measurement, Raw Score, Normative Sample, Percentile Rank, and National Quarters
4. Brief Description of the School or District in Terms of Relevant Descriptors (e.g., Sex, Race, Social Status)
5. Comparison of the Descriptors of the School or District with the Descriptors of the National Sample
6. Statement of Similarity of the Two Samples and the Necessity of Having Similar Samples When Comparing Student Test Performance
7. Graphical Representation of the Test Results by Grade Level and by Skill Area (The results should be presented by school and by district)
8. Comparison of this Year's Growth Results with the Results of the Previous Year, Stressing Signs of Growth
9. Comparison of this Year's Results for the School or District:
   A. What programs are in operation to correct known weaknesses? What programs are being considered?
   B. What programs work in those schools in the district that have higher test scores? Can ideas be taken from those schools and used in schools with lower test scores?

and curricula should be made based on the test results. Such suggestions can be incorporated into the last section of the document, if desirable.

The importance of informing school and district staff has been stated quite succinctly by the National School Public Relations Association (1976): "Peporters have been known to zero in immediately on a school whose scores were particularly high or low. The principal or teachers who shrug off the interview with an 'Oh, we're not worried about that Mickey Mouse program' will undercut whatever communication efforts are made by the central office" (p. 48). Once the document has been prepared and school and district staff informed, the next step is to inform the news media and the parents.

Informing the News Media

The first step in preparing for the news media is to write a news release. The focus of this release should be a summary of the test results. The release should be no more than two double-spaced legal size pages in length. An example of a news release is presented in Appendix A. After the preparation of the press release, reporters should be called in for a conference. The district test coordinator will probably be responsible for coordinating the activities related to the press conference. The purpose of the conference with the reporters is to explain the information presented in the press release and to supplement the verbal statement with graphs and tables where appropriate.

With respect to both the release and the conference, use positive statements whenever possible. That is, talk about strengths first, then weaknesses. Mention the percent of students above a certain level, not the percent below it.

In preparing this handbook certain terms have been avoided. Terms such as grade-equivalent scores and total battery scores should be avoided when discussing the test results with the news media and the public in general. Such terms should be avoided because of the misinterpretation often associated with them. For example, grade equivalent scores fail to take into consideration the appropriateness of the content areas for the students. To think that a third grade student whose third grade test performance indicates that he is "at the 6th grade level" and can, in fact, work mathematics problems contained in the 6th grade curriculum is inappropriate. The "6th grade level" only means that he works third grade mathematics problems extremely well for a third grade student. Rather than explain this important issue to every audience, the suggestion is that grade-equivalent scores not be presented.

Finally, the National School Public Relations Association suggests following a few basic rules in dealing with the news media. These rules are as follows:

> *Give the media everything you have. Don't hide anything. They'll probably find out anyway, and then you'll be in real trouble.
>
> *Don't belittle or downgrade any of the data.
>
> *Prepare charts or tables to hand out. Without distorting the data, you may be able to arrange the statistics in a more easily understood format.
>
> *Don't let the reporters get you on the defensive. Once they do, it's downhill the rest of the way. (p. 51)

At the end of the press conference, inform the media that the district

test coordinator has been charged with quickly and accurately responding to additional questions and inquiries from the press and the public. If your summary is well written and the above rules are followed, your relationship with the news media will likely be positive and productive.

Informing Parents

Informing parents is a continuous activity. Parents should be informed throughout the school year of the testing program and the test results. At the beginning of the school year, parents should be informed of the nature and purpose of the testing program. Then, about one month before the scheduled administration of the tests, parents should be informed of the pending administration. After the tests have been administered, the first information that parents receive will quite likely come from their children and the press. If the press has been properly informed, the first reaction by the parents will likely be positive. Furthermore, if the test results are communicated to the students, additional benefits may accrue. As the National School Public Relations Association writes, "Think how much a parent's mind would be eased if 12-year-old Jimmy would come home and say, 'We got our test scores back today and our teacher said we did pretty well except in arithmetic so we worked on a lot of problems and I think I understand it better.' Scratch one worried mother." (p. 54).

Communicating test results to individual parents should be done in a parent-teacher or parent-guidance counselor conference. Before the meeting with the parents, the Individual Test Record for the child should be re-

trieved from the child's permanent record file. Begin the conference by explaining the meaning of the points and "X's" on the graph. Then, discuss the child's performance on each test. Again, as in the case of the news media, focus first on the strengths of the child, then on the weaknesses. Indicate what the school intends to do about the weaknesses, as well as what the parents can do. At the end of the conference, summarize what has been said in order to minimize misunderstandings on the part of the parents.

Providing Follow-Ups

The surest way of destroying the effectiveness of a testing program is to view the program as a once-a-year thing, like paying taxes on April 15th. This point has been alluded to earlier in the discussion about ways of informing parents. The testing program should be seen and explained as an integral part of the district's educational program. If this is true, opportunities need to be provided for teachers, news media, and parents to follow-up the results of the testing program.

Suzanne Stemnock (1974) suggests several ideas for follow-ups that are worthy of consideration by district personnel.

First, briefings can be provided for members of the news media who want additional information. These briefings could focus on a more detailed discussion of various components of the testing program or on supplementary test results.

Second, print a booklet for widespread distribution. This booklet would contain all press articles and a commentary by the district test

coordinator.

Third, following the release of the test results, a panel of administrators may review the tests to identify strengths and weaknesses of students at each grade level tested. This review should focus on the test results by specific content category rather than by general skill areas and/or subskill areas which are much too global. Thus, questions should be raised such as How did our third grade students perform in the area of syntactical relationships (one of the content categories of the language test)?; Should we include syntactical relationships in our third grade curriculum (since we do not at the present time)?; Should we emphasize syntactical relationships more than we do?; Should we devise a new instructional program to teach students syntactical relationships?.

Fourth, reading, language, mathematics, reference skills, science, and social studies supervisors can visit each school to discuss changes that can be made in hopes of improving student achievements in the upcoming year. As much as possible, suggestions for changes should come from the staff of the school, not from the supervisor. Teachers tend to reject "improvements" that are suggested by "outsiders".

Finally, according to Stemnock, workshops can be provided for parents. The focus of these workshops should be on ways in which parents can help to reinforce the school's programs in their homes. Suggestions for home activities that likely will aid children's learning can be made.

In general, then, follow-ups after the release of the test results are important if full use of the results is to be made. The test results <u>should not</u> be viewed as "one-shot deals", having little or no relationship

to the educational program of the schools. Rather, test results can provide useful information for improving the educational program. If this is to be done, all persons concerned (that is, parents, teachers, and reporters) must be informed of the testing program, the results of the testing program, the relationship of the testing program to the instructional program, and programs for improving the weaknesses identified by the test results. Follow-ups such as those mentioned by Stemnock help to convince people of the role of testing in the total educational program of the school or district.

REFERENCES

Albohm, J. Releasing test scores: urgent or unthinkable? School Administrator. May 1973, pp. 19-20.

Florida State Department of Education. Florida Statewide Assessment Program: School Predicted Scores, 1975-76, (Statistical Report Series No. 77-07). Tallahassee, Florida, Florida State Department of Education, 1977. (ERIC Document Reproduction Service No. ED 142 588).

Hawes, G. R. Releasing test scores: urgent or unthinkable. Nation's Schools 1972, 89, 41-55.

Hutcheson, L. D., & Schabacker, W. H. Benchmarks - Utilization Guide for Georgia Statewide Test Scores. Atlanta, Georgia: Georgia State Department of Education, 1975. (ERIC Document Reproduction Service No. ED 117 183).

Larsen, E. P. Opening Instructional Ledger Books - A Challenge to Educational Leadership: Suggestions for Talking to School - Community Groups about Testing and Test Results. (RM Report No. 28) Princeton, New Jersey: Eric Clearinghouse on Tests, Measurement and Evaluation, 1974. (ERIC Document Reproduction Service No. ED 009 425.)

National School Public Relations Association. Releasing Test Scores: Educational Assessment Programs, How to Tell the Public. Arlington, Virginia: National School Public Relations Association, 1976. (ERIC Document Reproduction Service No. ED 119 322).

Rogers, W. T., & D'Angelo, G. A. Parents, teachers, and school administrators as targets for statewide assessment information. A paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, California, April 1976.

REFERENCES

Rudman, H. C. The standardized test flap. Phi Delta Kappan, 1977 59(3), 179-185.

Schramm, C. F. Comparing schools by academic achievement scores. Phi Delta Kappan, 1976, 58(4) 324-326.

Stemnock, S. Releasing Standardized Achievement Test Scores to the Public. Washington, D.C.: Educational Research Service, 1974 (ERIC Document Reproduction Service No. ED 106 306).

Thorndike, R. L. & Hagen, E. P. Presenting the test results to the public: In Measurement and Evaluation in Psychology and Education, (4th Edition), New York, New York: John Wiley and Sons, 1977.

Appendix A

An Example of a District News Release

FOR IMMEDIATE RELEASE
(date)

DISTRICTWIDE TEST
RESULTS ANNOUNCED

CITY -- Relative to the nation, third, sixth, and eleventh grade students in District #99 are achieving quite well in school. Summary results of the April testing were announced yesterday by District Superintendent A.T. Jones.

A total of 1,215 pupils in grades three, six, and eleven was administered the Comprehensive Tests of Basic Skills (CTBS) on April 16-18, 1979. The CTBS are nationally standardized achievement tests which assess student achievement in six general areas: reading, language, mathematics, reference skills, science, and social studies.

Jones said that District #99 students generally performed highest in the areas of language, science, social studies, and reference skills. (Reference skills refer to such abilities as using the dictionary and library to locate information.) In all three grades, more than 45 percent of District #99 students scored above the national average in these four areas. The lowest performance for all three grades was in mathematics.

In the area of language, 48 percent of District #99 third graders scored above the national average; 45 percent of the sixth graders scored above the national average; and 41 percent of the eleventh graders scored

above the national average.

In reading, 44 per cent, 39 percent, and 36 percent of third, sixth, and eleventh grade students, respectively, scored above the national average.

In mathematics, 40 percent of the third grade students scored above the national average. The percentages in grades six and eleven were 33 and 32, respectively.

The highest average scores for District #99 students were in the area of reference skills. Fifty percent of the third graders, 46 percent of the sixth graders, and 42 percent of the eleventh grade students scored above the national average in reference skills.

The results in science and social studies were quite similar. Approximately 47 percent of the third grade students scored above the national average in these two areas. Approximately 42 and 40 percent of the sixth and eleventh grade students, respectively, scored above the national average in science and social studies.

At the release of the summary results, Jones urged caution in interpreting the results. Since the students in the national sample differ considerably from the District #99 students in terms of socio-economic status, racial composition, and other characteristics, the scores of the district sample and the national sample are likely to be somewhat different. If one were to compare the top students in District #99 with the top students in the national sample at each grade level, for example, the differences in test performance would be slight. The bottom students in District #99, however, tend to score much lower than the bottom students in the national sample. These bottom students in the district tend to come

from more culturally deprived homes than their counterparts in the national sample.

Since the scores were lowest in reading and mathematics, a panel of teachers and administrators has been formed to study this problem. The panel will begin by comparing the areas of emphasis of the test with the areas of emphasis of the reading and mathematics programs used in the district. If areas are found that should be emphasized in the programs but currently are not being emphasized, changes in the programs will be considered. If changes are proposed, parents will be asked to comment on the changes prior to instituting such changes.

Parents interested in obtaining additional information about the test results are encouraged to contact Mrs. Flora Smith, the District Test Coordinator. Mrs. Smith can be reached at 555-1234.

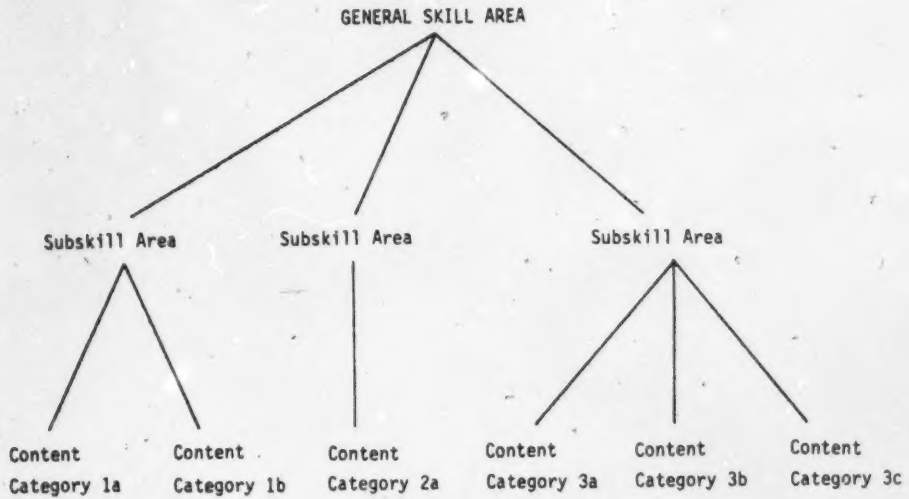


Figure 1. Interrelationships among General Skill Area, Subskill Areas, and Content Categories.

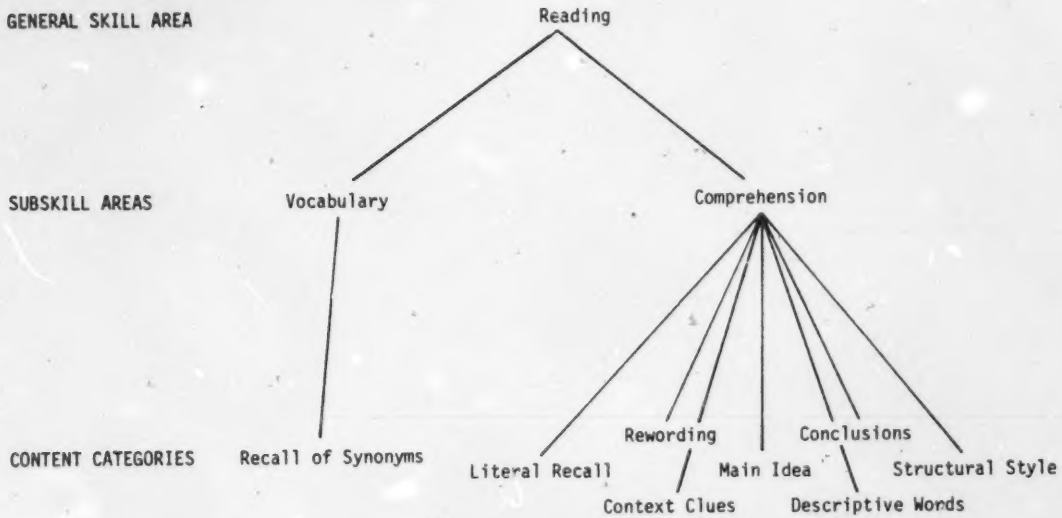


Figure 2. Interrelationships among General Skill Area, Subskill Areas, and Content Categories for Reading.

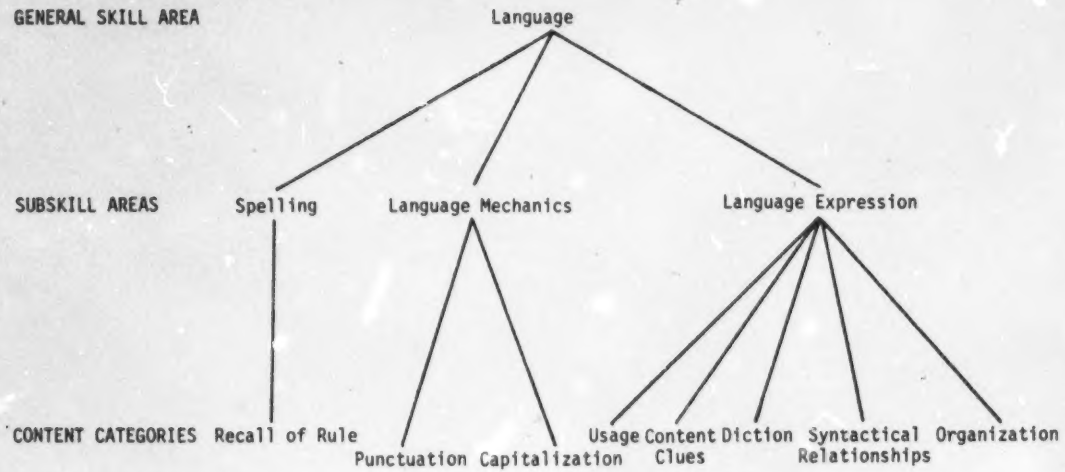


Figure 3. Interrelationships among General Skill Area, Subskill Areas, and Content Categories for Language.

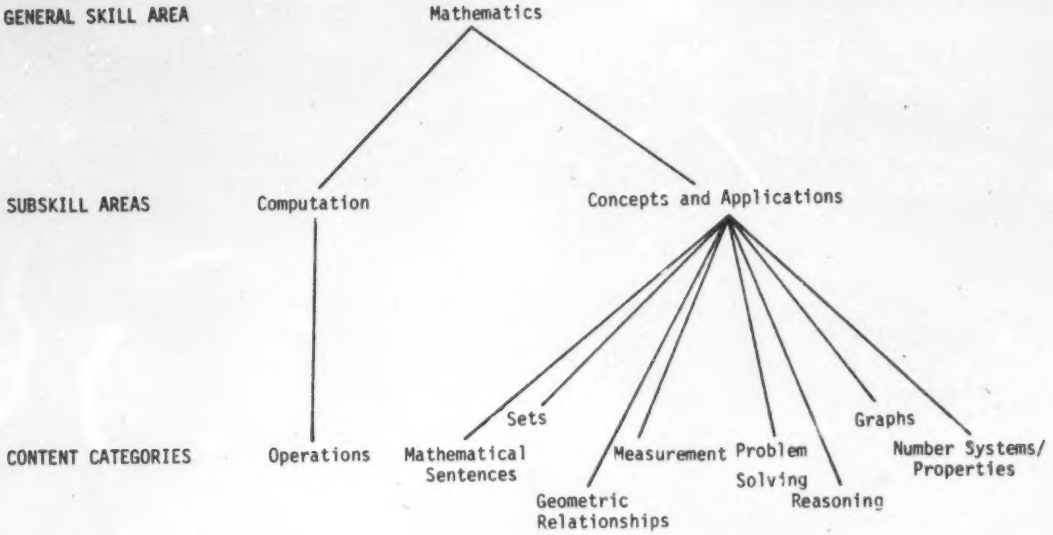


Figure 4. Interrelationships among General Skill Area, Subskill Areas, and Content Categories for Mathematics.

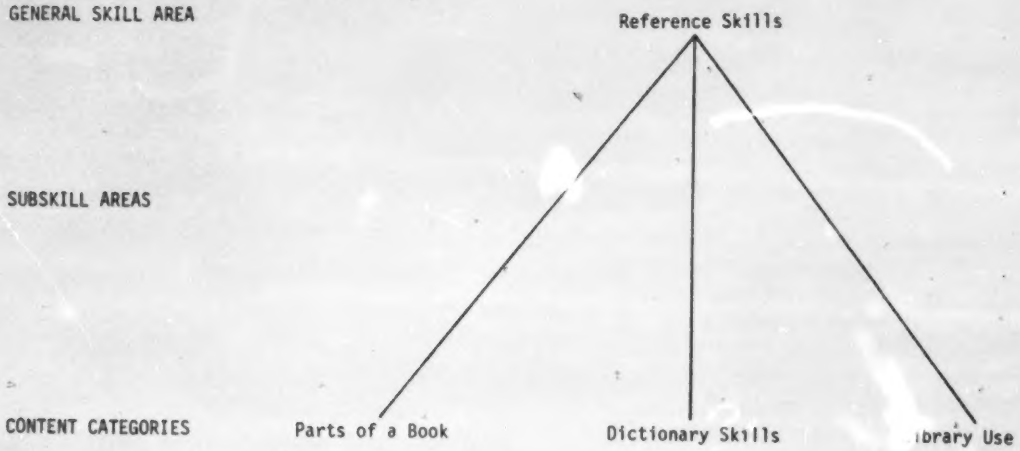


Figure 5. Interrelationships among General Skill Area, Subskill Areas, and Content Categories for Reference Skills.

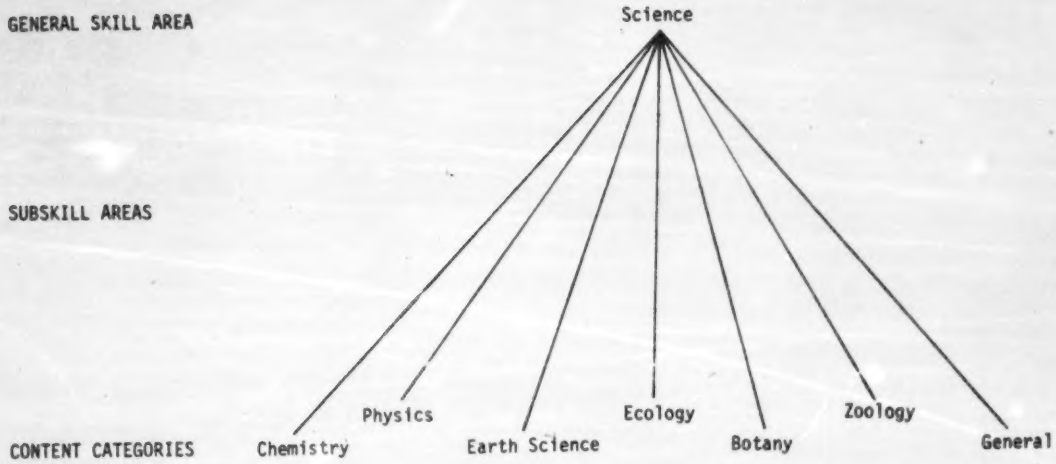


Figure 6. Interrelationships among General Skill Area, Subskill Areas, and Content Categories for Science.

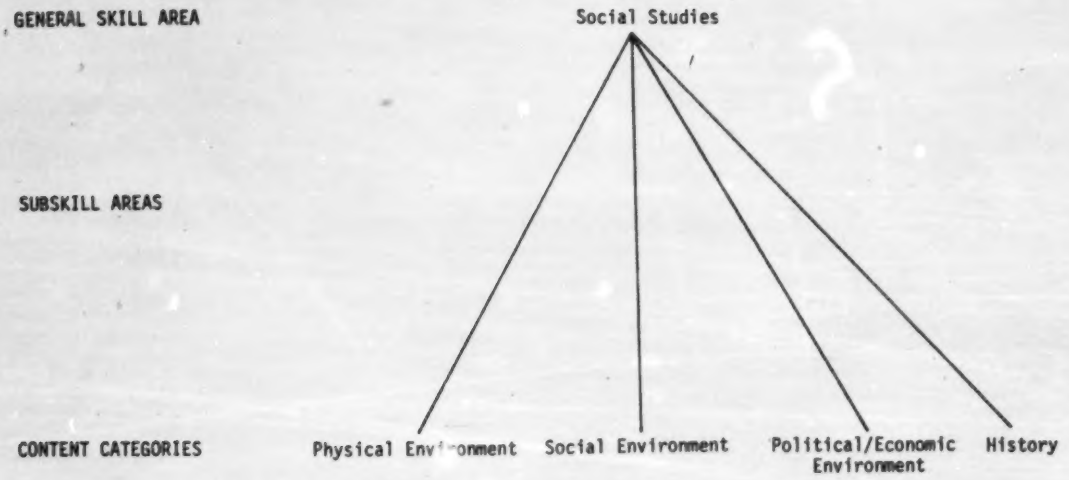


Figure 7. Interrelationships among General Skill Area, Subskill Areas, and Content Categories for Social Studies.

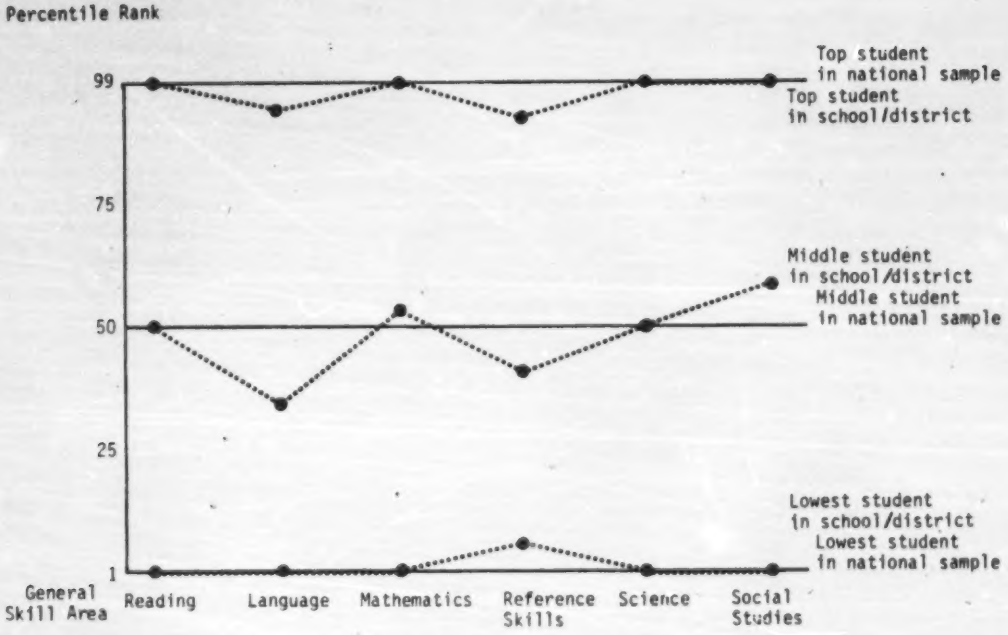


Figure 8. A Summary of School/District Performance on the CTBS/S (by General Skill Area).

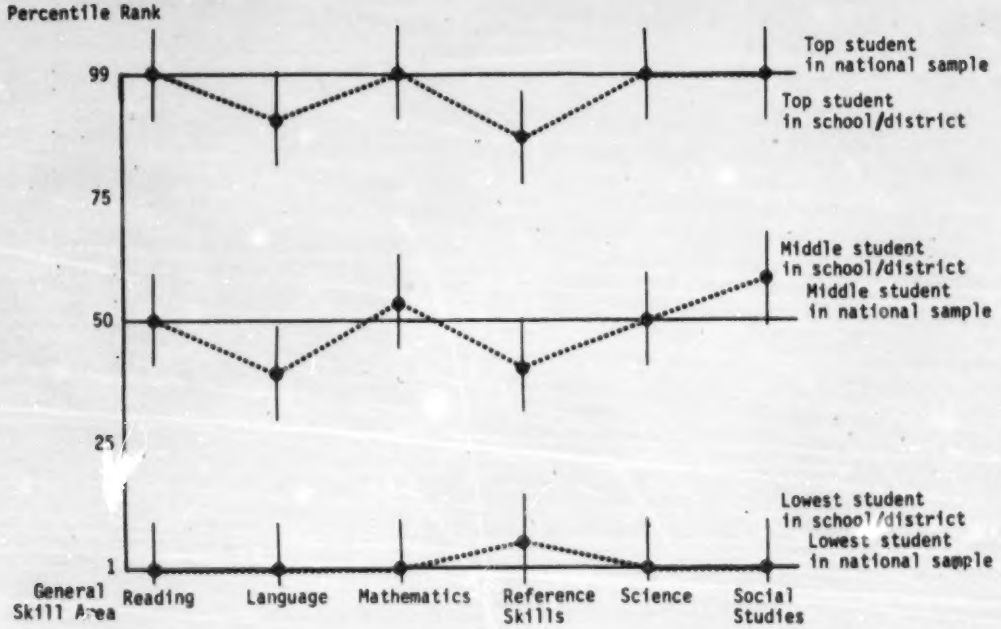
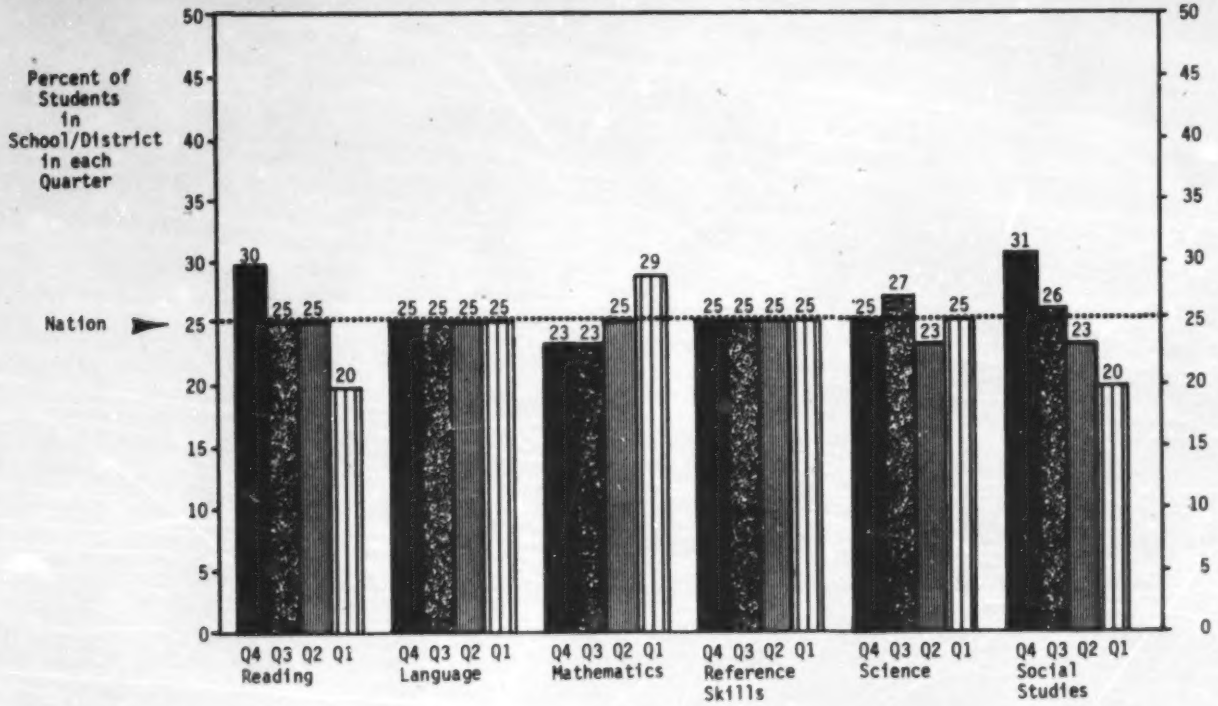


Figure 9. A Summary of School/District Performance on the CTBS/S (emphasizing errors of measurement) (by General Skill Area).



Q4 = upper quarter, percentile ranks 76-99
Q3 = upper middle quarter, percentile ranks 51-75
Q2 = lower middle quarter, percentile ranks 26-50
Q1 = lower quarter, percentile ranks 1-25

Figure 10. Graphical Display of School/District Results on the CTBS/S (by General Skill Area).